简版健康素养量表的开发：基于经典测量理论和项目反应理论

孙小楠 [1]　陈珂 [2]　武运筹 [3]　汤靖琪 [4]　王飞 [5]　孙昕霙 [6]　贺苗 [1*]　吴一波 [6*]

[1]（哈尔滨医科大学人文社会科学学院 哈尔滨 150081）

[2]（安徽大学哲学学院心理学系 合肥 230039）

[3]（西南大学心理学部 重庆 400715）

[4]（华东师范大学心理与认知科学学院，上海，200062）

[5]（北京师范大学心理学部认知与学习国家重点实验室 北京 100875）

[6]（北京大学公共卫生学院 北京 100191）

*通讯作者：贺苗，教授；hemiao767@163.com;吴一波，博士研究生；bjmuwuyibo@outlook.com

摘要：

[目的] 简化健康素养量表并在中国人群中进行心理测量学检验。

[方法] 纳入居民 7449 人进行量表的测评，并将数据随机生成 2 个数据集，进行描述性分析，随后应用经典测量理论（CTT）以及项目反应理论（IRT）中的 Mokken 模型对原条目进行筛选，并对精简后的量表进行信效度等的验证性分析。

[结果] 使用经典测量理论和 Mokken 模型分别简化出一个 9 条目版本的量表（HLS-SF9）和一个 4 条目版本的量表（HLS-SF4）。简化后的健康素养量表 HLS-SF9 与 HLS-SF4 均无天花板效应、地板效应，二者的 Cronbach’s α 系数为 0.913 和 0.842，HLS-SF4 经探索性因素分析(EFA)检验出一个公因子，HLS-SF9 经验证性因素分析(CFA)结果显示其三维模型的各适配指标均优秀，以领悟社会支持量表与家庭健康量表为校标量表，HLS-SF9 与 HLS-SF4 均有显著的正相关（r=0.367， p<0.001；r=0.340， p<0.001）（r=0.292， p<0.001；r=0.237，p<0.001），表明实证效度良好。同时 HLS-SF9、HLS-SF4 与原量表所测量的内容有高度一致性。

[结论] 简化后的健康素养量表具有良好的信效度，是能够快速测评中国人群健康素养的有效工具。

关键词：健康素养； 条目筛选； 经典测量理论； Mokken 模型

Development of a Short Version of the Health Literacy Scale Short-Form:Based on Classical Test Theory and Item Response Theory

Abstract:

[Objective] Simplify health literacy scales and conduct psychometric tests in Chinese cohorts.

[Methods] A total of 7449 residents were included in the evaluation of the scale, and the data were randomly generated into 2 data sets for descriptive analysis, validation factor analysis, application of classical measurement theory (CTT) and Mokken model in item response theory (IRT) for entry screening, and validation analysis of the streamlined scale for reliability and validity, etc.

[Results] A 9-item version of the scale (HLS-SF9) and a 4-item version of the scale (HLS-SF4) were simplified using CTT and the Mokken model, respectively.The Cronbach’s α coefficients of HLS-SF9 and HLS-SF4 were 0.913 and 0.842, HLS-SF4 was tested for one common factor by exploratory factor analysis(EFA), and the results of the confirmatory factor analysis(CFA) of HLS-SF9 showed that all the fitness indicators of its three-dimension model were excellent. And there was a significant positive correlation between the Perceived Social Support Scale and the Family Health Scale Short-Form as the calibration scale (r=0.367, p<0.001;r=0.340, p<0.001) (r=0.292, p<0.001; r=0.237, p<0.001), indicating good empirical validity. At the same time, HLS-SF9 and HLS-SF4 are highly consistent with the content measured by the original version of the scale.

[Conclusions] The simplified Health Literacy Scales have good reliability and validity, and are reliable and effective tools for quickly assessing the health literacy of Chinese people.

Keywords: Health Literacy Scale Simplification Classical Test Theory Mokken Model

健康素养系个体获取、理解健康信息以保障、推进自身健康建设的能力[1]。世界卫生组织强调，健康素养作为一种认知能力和社会技能水平的象征[2]，是健康的重要决定因素[3]。健康素养水平的限制易让公众难以完整地认识、了解疾病，难以实现对医疗资源的高效、充分利用，特别是影响慢性病患者的自我疾病管理，从而存在较差的服药依从性，甚至增加住院率与死亡率[4-6]。我国《“健康中国2030”规划纲要》强调居民的健康素养情况是战略目标的核心内容。提升健康素养水平应成为健康教育、患者管理和健康促进等方面的重要构成[7]，这就要求个人要充分了解并使用健康信息，以便充分管理自身的健康问题，降低社会成本。因此，一个能客观、全面且准确测评健康素养的工具是开展此类研究的前提。

随着健康素养领域得到国内外学者的关注，用于测量公众的健康素养工具研发活跃,目前常用成人功能健康素养测试（test of functional health literacy in adult，

TOFHLA）、欧洲健康素养调查（the European health literacy survey questionnaire，HLS-EU-Q）等问卷[8-9]。结合以上研究，Tuyen V. Duong 在 HLS-EU-Q47 的基础上，编制了适用于评估亚洲国家公众健康素养的 12 条目健康素养量表（short-form health literacy survey questionnaire，HLS-SF12），以良好的信效度支持了 HLS-EU-Q47 的原始架构[10]。我国学者多使用国家卫健委制定的“全国居民健康素养监测调查问卷”，随着对健康素养研究的深入，陆续有学者研发了针对特殊人群的健康素养评估工具[11-13]。

目前课题研究中多倾向于多维度、多条目的测评工具，力争全面化评估受测者的临床或心理特质，但随之应用，冗长的工具也出现了一定的不足与缺陷：问卷中题目过多，使得受访者作答时间较长，作答耐心降低，作答认真度下降，问卷的真实性可靠性难以保障；同时易使受访者产生隐私受侵的心理，增加受访者的心理负担[14]。而简短版量表则能较大程度上的规避以上弊端，同时缩减问卷填写时间，利于推广受访人群与应用领域。此外，简短的健康素养评估工具可被纳入患者就诊评估等评估问卷，快速筛查出健康素养有限的群体，便于实施针对性的健康教育，评估干预效果[8][15]。

考虑到我国在健康素养领域缺失简便的测评工具，本研究尝试对 12 条目健康素养量表进行简化研究，这将有助于在更大规模的人群或临床环境中对健康素养进行简单而准确的评估，为今后的干预研究提供参考依据。

1 对象与方法

1.1 受测对象

于 2021 年 7 月 10 日-2021 年 9 月 15 日开展此次调查，采用多阶段抽样的方法，将中国 23 个省和 5 个自治区的省会、4 个直辖市直接纳入，并用随机数表法在每个省、自治区的非省会地级行政区中各抽取 2-6 个城市，共 120 个城市。每个城市至少招募 1 位调查员或 1 支调查团队。调查员需基于“2021 年第七次全国人口普查结果”的数据结果，使所获得样本的性别、年龄、城乡分布基本符合人口特征。纳入标准为：①年龄≥12 岁；②具有中华人民共和国国籍；③中国常住人口（年外出时间≤1 个月）；④自愿参加研究，填写知情同意书；⑤可自行完成网络问卷调查或在调查员帮助下完成问卷调查；⑥了解问卷每个条目所表达的含义。排除标准：①神志不清、精神异常者；②正在参加其他类似研究课题者；③不愿合作者。问卷回收后由两人背靠背进行逻辑检查和数据筛选。本研究已通过伦理审查（JNUKY-2021-018）。调查居民共 11668 例，回收有效问卷 11031 份，问卷有效率为 94.54%。从“中国家庭健康指数调查”数据中选择 18 岁及以上人群作为本研究的受测对象，最终纳入 7449 份数据，并随机分成两个样本集，其中样本集 1 共 3680 份，样本集 2 共 3769 份。

1.2 研究工具

1.2.1 一般资料问卷

由研究者编制，内容包括调查对象的性别、年龄、民族、户口类型、居住地类型、最高学历情况、婚姻状况、家庭人均月收入等。

1.2.2 健康素养简表

Tuyen V. Duong 等编制[10]的适用于公众健康素养测量的 HLS-SF12 分为 3 个维度，分别是卫生保健、疾病预防、健康促进，量表共 12 个条目，采用 4 级评分（1=非常困难，2=困难，3=容易，4=非常容易），使用公式计算标准化 HL 指数，指数范围为 0~50，指数越高代表健康素养水平越高。计算公式为，指数=（平均值−1）×（50/3），其中平均值是每个个体所有参与项目的平均值，1 是平均值的最小可能值（此时指数的最小值为 0），3 是平均值的范围，50 是指数的最大值。Tuyen V. Duong 报告健康素养简表的 Cronbach's α ＞0.70，卫生保健分量表的 Cronbach’s α 系数为 0.49~0.72，疾病预防分量表的 Cronbach’s α 系数为 0.64~0.77，健康促进分量表的 Cronbach’s α 系数为 0.59~0.81，内部一致性指标良好。施测时用的是汉化后的 HLS-SF12 中文版。本研究中该量表在数据集 1 的 Cronbach’s α =0.932，数据集 2 的 Cronbach’s α =0.933，总数据集的 Cronbach’s α =0.932。

1.2.3 领悟社会支持量表

由 Zimet 等[16]开发、姜乾金修订的领悟社会支持量表（Perceived Social Support Scale, PSSS）分为家庭支持、朋友支持和他人支持三个维度，每个维度含 4 个条目，共 12 个条目。量表选项从“极不同意”到“极同意”赋分为 1~7 分，得分越高领悟到的社会支持越丰富。本研究中该量表在数据集 2 的 Cronbach’s α =0.947，总数据集的 Cronbach’s α =0.948。

1.2.4 家庭健康量表

Crandall 和 Weiss-Laxer 等[17]编制，用于测评家庭健康功能，由王飞、武运筹等[18]汉化翻译。该量表分 4 个维度，共 10 个条目。题项为“非常不同意”~“非常同意”（1~5 分），其中第 6、9、10 题为反向计分。总量表得分越高表示家庭健康水平越好。本研究中该量表在数据集 2 的 Cronbach's α =0.845，总数据集的 Cronbach's α =0.846。

1.3 统计学分析

研究采用 SPSS 24.0、AMOS 24.0 和 R 4.2.1 进行数据处理，使用了描述性统计分析、验证性因子分析、经典测量理论精简条目、Mokken 模型精简条目、验证性分析等分析方法。为了保障简版量表的有效性，以及避免出现样本量不足造成的研究能力降低的情况，本研究对研究需要的最低样本量进行了计算[19]：假设使类内一致性水平（ICC）达到 0.90，I型错误概率α为 0.05，此时实现 95%的统计功效需要 223 名被试。本研究收集了 7449 份有效数据（包括条目筛选和验证分析两份数据集在内），说明样本量足以进行后续数据分析。

研究首先对数据集1和数据集2这两个数据集的社会人口学信息进行描述性统计，显示变量各分类的人数及百分比。

通过经典测量理论在项目分析时最常用的四种方法：项目间残差相关法、相关系数法、CITC法、独立样本 *t* 检验法对原量表的每个条目进行分析。项目间残差相关法计算各项目与其余项目残差相关之和，表现为项目残差的MI值相加，若MI值之和越小，残差对本维度的解释越弱，条目本身对维度的解释越强[21]。相关系数法是计算各条目与量表总得分的皮尔逊相关系数，选取量表中单个条目得分与量表总分的相关系数的绝对值较大的、且存在显著的统计意义的条目[22]。项目-总体相关系数（CITC）法根据量表的内部一致性筛选条目，通过计算总量表或单个维度的克隆巴赫系数(Cronbach's α)，比较删除某一条目后Cronbach's α的变化。如果某条目去掉后总量表或单个维度的 α 系数有明显升高，表明应当删除，因为该条目的存在会降低量表或维度的内部一致性，反之则保留[23]。独立样本 *t* 检验法首先将量表总分由高到低排列，高分组为总分最高的27%，低分组为总分最低的27%，随后进行独立样本 *t* 检验，若结果显示高低分两组被试在某条目上平均得分不存在显著性差异，则应当删除该条目[24]。

通过项目反应理论中的Mokken模型[25]进行分析。Mokken模型属于非参数项目反应理论，与参数项目反应理论相比，它提出更适应实际情景、更有弹性的框架，还更适宜短量表使用[26]。利用R语言的“Mokken”包对健康素养量表全量表进行分析，包括对所有条目的单维性检验、局部独立性检验、单调性检验。利用自动项目选择算法（Automated Item Selection Procedure，AISP）检验量表单维性，AISP实施时，从c=0开始，到c=0.55结束，步长0.05。当c值越大时，如果测验是单维的，可能会出现以下三个阶段：绝大部分或全部条目合并为一个量表；形成一个容量较小的量表；形成一个或几个小量表，同时许多项目被删除[27]。同时，可以利用同质性系数(homogeneity coefficients)检查条目设置是否合理。同质性系数分为3类：项目对(H_{ij})、项目(H_i)和量表(H_s)。同质性系数H值的值越高，测验所得总分对被试潜在特质的排序越准确，Mokken依据自身经验，认为H_{ij}必须大于0，H_i和H不能低于0.3。更具体的H值规定为：当H＜0.3时，量表不合格；当0.3≤H＜0.4时，量表的准确程度较弱；当0.4≤H＜0.5时，量表的准确程度中等；当H≥0.5时，量表的准确程度强[28]。局部独立性检验利用条件关联程序完成，由三个条件关联指数$W^{(1)}$、$W^{(2)}$、$W^{(3)}$检验，被标记的项目要逐一删除，删除的原则是：具有最多 W 标志的项目被删除，直到只剩下没有标志的项目。如果项目具有相同数量的标志，则同质性系数H_i更小的项目将被删除[29]。最后，检验每个项目的单调性。单调性将最小紊乱系数（#vi/#ac）、显著性（#zsig）和Crit的数值作为评价标准，当以上三个指标等于0时，说明符合单调性假设。但在实际应用中，最小紊乱系数小于0.3可接受[30]，显著性小于1.96可接受[31]。当Crit＞80时，违背单调性假设；当40≤Crit≤80时，应按照条目内容和量表使用目的考虑是否删除；当Crit＜40时，则可认为该条目基本满足单调性，个别违反单调性假设的情况可以看作是被试抽样误差所致[32]。

基于数据集2的数据进行验证性分析。在探索性分析的基础上，对得到的简版量表进行验证性分析，分为天花板和地板效应检验、信度检验和效度检验简版量表。地板和天花板效应分别反映了得分最低和最高的参与者的反应，建议最低或最高水平的百分比为15%或更低。如果超过15%的受访者分别获得了可能的

最低或最高分数，则认为存在地板或天花板效应[33]。若存在天花板效应，则该量表在实际使用中，由于顶端水平上选择数量增多从而导致后续数据分析中各指标的有效性受到影响。地板效应与之相反。信度指标若均＞0.7 表明可接受[34]。效度检验分为结构效度分析和实证效度分析。结构效度检验中，HLS-SF4 采用探索性因素分析，HLS-SF9 则采用验证性因素分析。然后通过比较 12 条目的原量表和开发的简版量表与 2 个相关概念的相关性，进行实证效度检验。根据数据类型，使用 Pearson's r 计算相关性。为检验简版量表与完整版量表所测内容的一致性程度，我们利用 RStudio 中的“lpSolve”和“irr”包计算了类内相关系数（Intra-class correlation coefficient,ICC），ICC 可以反映测量之间的相关程度和一致性。ICC 的评价标准为：当 ICC＜0.50 时，被解释为一致性差；当 0.50≤ICC＜0.74 时，被解释为一致性中等；当 0.75≤ICC≤0.90 时，被解释为一致性好；ICC＞0.90 时，被解释为一致性优秀[35]。

2 结果

2.1 受测对象的社会人口学信息

在样本数据集 1 的 3680 名受访者中，男性 1608 名（43.7%），汉族 3449 名（93.7%），常住城镇者 2700 名（73.4%），农业户口 1524 名（41.4%）。在数据集 2 的 3769 名受访者中，男性 1678 名（44.5%），汉族 3544 名（94%），常住城镇者 2170 名（57.6%），农业户口 1599 名（42.4%）。完整的人口统计细节见表 1。

表 1 受测对象的一般人口学特征

社会人口学信息	数据集 1		数据集 2	
	人数	百分比(%)	人数	百分比(%)
性别				
男	1608	43.7	1678	44.5
女	2072	56.3	2091	55.5
民族				
汉族	3449	93.7	3544	94
少数民族	231	6.3	225	6
常住地				
城镇	2700	73.4	2749	72.9
农村	980	26.6	1020	27.1
户口性质				
非农业	2156	58.6	2170	57.6
农业	1524	41.4	1599	42.4
是否负债				
是	1514	41.1	1553	41.2
否	2166	58.9	2216	58.8
家庭人均月收入				
≤3000	1113	30.2	1118	29.7

3001-6000	1442	39.2	1476	39.2
6001-9000	588	16	610	16.2
≥9001	537	14.6	565	15
年龄段				
≤30	1333	36.2	1379	36.6
31-45	1067	29	1100	29.2
46-60	873	23.7	866	23
≥61	407	11.1	424	11.2
最高文化程度				
未接受正规学历教育/小学	353	9.6	387	10.3
初中/中专	658	17.9	654	17.4
高中	370	10.1	417	11.1
大专/大学本科	2058	55.9	2068	54.9
硕士研究生/博士研究生	241	6.5	243	6.4
婚姻状况				
未婚	1200	32.6	1239	32.9
已婚	2329	63.3	2330	61.8
离异	56	1.5	95	2.5
丧偶	95	2.6	105	2.8
职业状态				
学生	812	22.1	864	22.9
在职	1776	48.3	1781	47.3
退休	318	8.6	333	8.8
无固定职业	774	21	791	21

2.2 完整版量表的心理测量学检验

基于数据集 1 的数据，对 HLS-SF12 的心理测量学指标进行验证性因子分析，如图 1 所示。结果显示 *NFI*=0.960，*GFI*=0.957，*AGFI*=0.934，*CFI*=0.962，*RMSEA*=0.068，各拟合指标表明原量表模型拟合良好。计算内部一致性系数得出 Cronbach's α=0.918，量表信度良好。

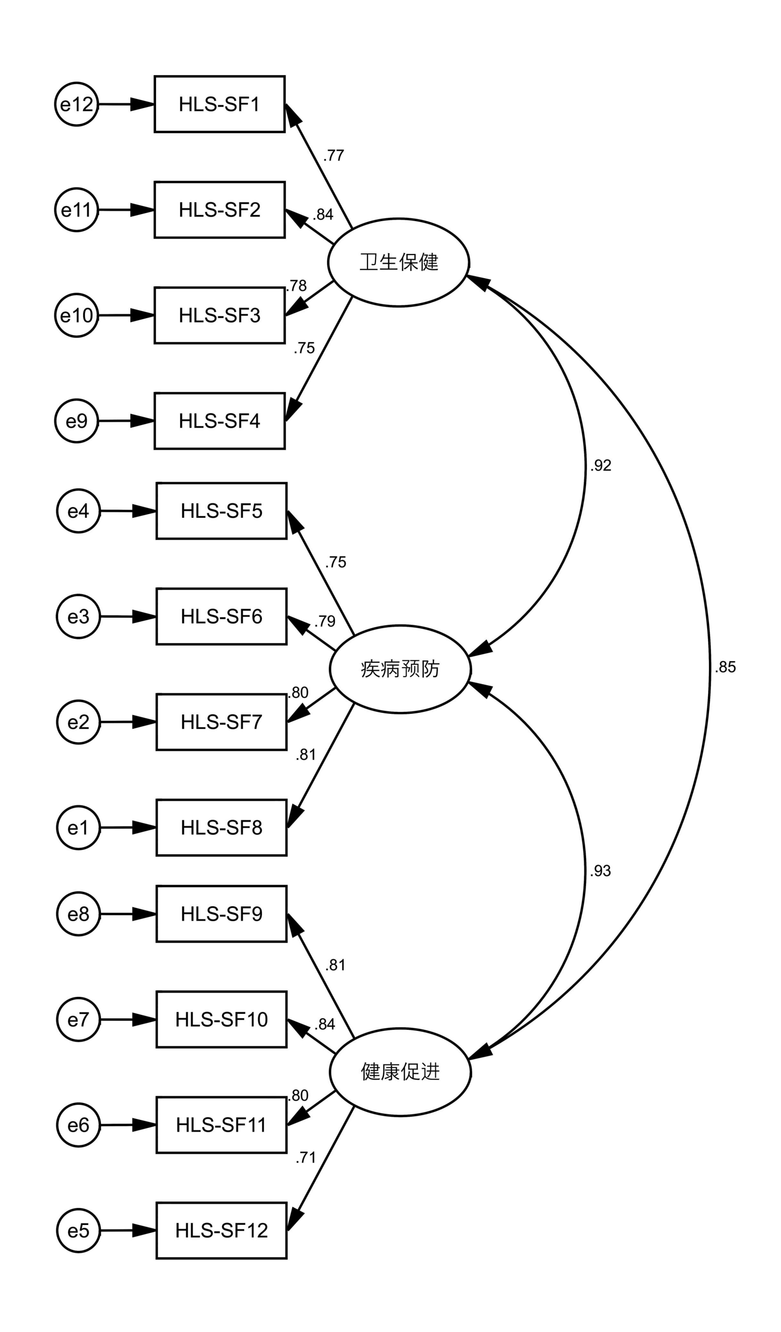


图 1 完整版量表 HLS-SF12 的验证性因子分析

2.3 探索性分析

2.3.1 基于经典测量理论的条目精简

首先，采用项目间残差相关法进行项目分析，MI 值的门槛值使用默认值。结果显示，各维度中残差 MI 值最大的条目分别为条目 3、条目 5 和条目 11，说明这几个条目对本维度的解释力在每个维度的所有条目中是最小的，故考虑剔除。

其次，采用相关系数法进行项目分析。结果表明，完整版健康素养量表与各条目之间的相关系数均高于 0.71（r=0.716~0.797），说明这些条目与量表的一致性良好，均考虑保留。

第三，采用项目-总体相关系数法（Corrected Item-Total Correlation,CITC）对

量表进行分析，发现删除每项后的 Cronbach's α系数在 0.924~0.928 之间，删除条目后内部一致性系数均有所下降，因此量表中没有条目需要删除。

最后，以健康素养量表 HL 指数最高的 27%(≥37.5 分)和最低的 27%(≤30.556 分）划分高分组与低分组，进一步做独立样本 *t* 检验，结果显示，量表的高、低分组在各条目上的得分均存在显著性差异(p<0.001)，因此量表中没有对应的条目需要删除。

综上所述，运用四种经典测量理论的方法对 HLS-SF12 进行精简，结果表明条目 3、条目 5 和条目 11 在项目间残差相关法中考虑删除，因此将这 3 个条目删除，形成一个 9 条目的三维简版量表。基于经典测量理论简化后的健康素养量表(HLS-SF9)共包括 9 个条目：条目 1、条目 2、条目 4、条目 6、条目 7、条目 8、条目 9、条目 10 和条目 12。具体条目分析结果如表 2 所示。

表 2　基于经典测量理论的 5 种条目分析方法结果汇总

条目	项目间残差相关法	相关系数	CITC	极端值法（*p* 值）	被入选次数	删除条目
HLS1	215.869	0.730	0.928	＜0.001	4	
HLS2	129.501	0.791	0.925	＜0.001	4	
HLS3	**517.124**	0.745	0.927	＜0.001	3	√
HLS4	230.189	0.725	0.927	＜0.001	4	
HLS5	**361.478**	0.737	0.927	＜0.001	3	√
HLS6	210.054	0.760	0.926	＜0.001	4	
HLS7	18.808	0.785	0.925	＜0.001	4	
HLS8	123.271	0.785	0.925	＜0.001	4	
HLS9	8.414	0.770	0.925	＜0.001	4	
HLS10	5.270	0.797	0.924	＜0.001	4	
HLS11	**41.051**	0.762	0.926	＜0.001	3	√
HLS12	5.978	0.716	0.928	＜0.001	4	

2.3.2 基于 Mokken 模型的条目精简

首先，对 HLS-SF12 进行 Mokken 模型分析。利用 AISP 检验量表单维性，从 c=0 开始，到 c=0.55 结束，步长设置为 0.05。结果表明，AISP 当 c 设置在 0~0.55 之间时，均只能得到一个维度，且所有项目均在该维度中。然后，计算量表的各同质性系数。结果表明，本研究中，H_{ij} 均大于 0.43，H_i 均大于 0.53，H=0.609，详情见表 3。这说明本研究使用同质性系数不能删除条目。

再进行局部独立性检验，即进行条件关联分析。第一轮分析中，指标 $W^{(1)}$ 表明第 11 项有 6 个标记，第 2 和 8 项各有 4 个标记，第 10 项有 1 个标记，故先将条目 11 删除。在之后的几轮分析中，根据指标 $W^{(1)}$、$W^{(2)}$ 和 $W^{(3)}$ 的结果，依次删除条目 8、条目 2、条目 12、条目 6、条目 9、条目 4、条目 10。经过条件关联分析，保留 4 个条目，分别是条目 1、条目 3、条目 5、条目 7。

随后进行单调性检验，条目 3、条目 4、条目 5、条目 6、条目 9、条目 11 和条目 12 出现违反单调性的情况，其中条目 3 的最小紊乱系数为 0.02，显著性为 1，Crit 值为 13，考虑保留；条目 4 的最小紊乱系数为 0.03，显著性为 1，Crit

值为 27，考虑删除。以此类推，条目 4、条目 6、条目 9、条目 11 可以删除，详见表 3。

综上所述，基于 Mokken 模型简化后的健康素养量表(HLS-SF4)共包括 4 个条目：条目 1、条目 3、条目 5、条目 7。

表 3　基于 Mokken 模型的条目分析结果

Item	H(SE)	#vi/#ac	#zsig	Crit
HLS1	0.571 (0.012)	0	0	0
HLS2	0.622 (0.011)	0	0	0
HLS3	0.630 (0.012)	0.02	1	13
HLS4	0.593 (0.013)	0.03	1	17
HLS5	0.611 (0.012)	0.02	1	19
HLS6	0.605 (0.012)	0.03	1	21
HLS7	0.618 (0.011)	0	0	0
HLS8	0.635 (0.012)	0	0	0
HLS9	0.610 (0.012)	0.03	1	23
HLS10	0.629 (0.011)	0	0	0
HLS11	0.633 (0.013)	0.03	1	29
HLS12	0.558 (0.013)	0.02	1	25

注：#ac 表示进行是否违反单调性假设的检验次数；#vi 表示违反了单调性假设的次数；#vi/#ac 表示最小紊乱系数，也称违反限度；# zsig 表示违反单调性的显著系数

2.4 验证性分析

2.4.1 天花板和地板效应检验

HLS-SF9 和 HLS-SF4 的可靠性可通过最小的地板/天花板效应而得到加强。表 4 显示其在数据集 2 中的天花板和地板效应，均未超过 15%，说明得分最低或最高的被试可以相互区分，利于信度的测量。

表 4 两个精简版量表的 HL 指数得分情况

	最小值	最大值	M±SD	地板效应（%）	天花板效应（%）
HLS-SF9	0	50	35.08±7.96	0.2%	9.2%
HLS-SF4	0	50	33.05±9.05	0.5%	10.2%

2.4.2 信度验证

我们使用数据集 2 来检验两个简化后的健康素养量表的信度，分析显示，HLS-SF9 和 HLS-SF4 的 Cronbach's α系数为 0.913 和 0.842、折半信度为 0.871 和 0.815，各条目删除后的信度均小于等于 0.910 和 0.810，信度分析指标良好。

2.4.3 效度检验

2.4.3.1 结构效度

在数据集 2 中对 HLS-SF4 进行 Bartlett 球形检验和 KMO 度量。HLS-SF4 的 Bartlett 球型检验值为 5915.883（$p<0.01$），KMO 度量为 0.807，可以进行因子分析。随后探索性因素分析提取出 1 个特征根大于 1 的因子，从 CTT 的角度验证了其单维性，累计方差贡献率为 67.813%，各条目的因子载荷量均大于 0.81。此外，其累计方差贡献率和各条目因子载荷均高于原量表，对所测变量的解释力更强。

对 HLS-SF9 的 9 个条目进行验证性因子分析（见图 2），HLS-SF9 的验证性因子分析模型适配指标的检验结果显示 *CMIN/df*=10.844、*GFI*=0.985、*AGFI*=0.971、*NFI*=0.986、*CFI*=0.987 和 *RMSEA*=0.051，除 *CMIN/df* 外均达到理想标准，需要说明的是卡方值与自由度的比值(CMIN/DF)消除了自由度的影响，但没有消除样本容量的影响，由于样本数量为 7449，属于大样本，而相关研究表明当样本数较大时，模型整体适配度的卡方值就会随着样本数增大而显著增大，这时只需要考虑其它重要指标，而这个指标就可以忽略。因此三维 9 条目的 HLS-SF9 模型拟合结果较好。

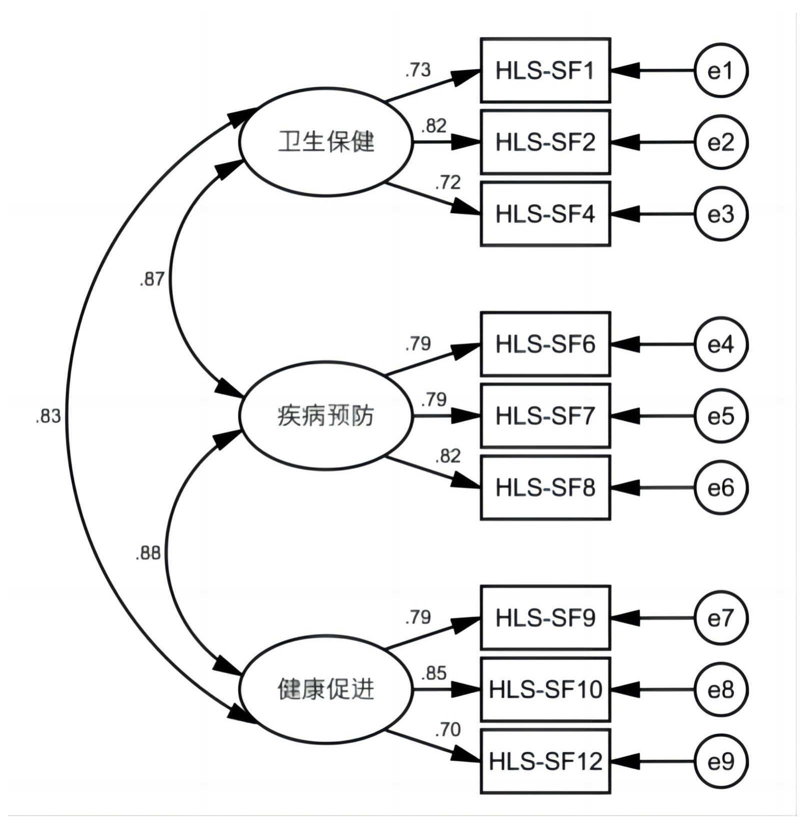


图 2 HLS-SF9 验证性因子分析模型

2.4.3.2 实证效度

以往研究表明，健康素养与领悟社会支持、家庭健康均呈显著相关[36-38]。我们运用数据集 2，将领悟社会支持、家庭健康和两个简化版健康素养量表同时进行相关分析，结果显示，HLS-SF9 和 HLS-SF4 与领悟社会支持有显著的正相关（r=0.367，p<0.001；r=0.292，p<0.001），与家庭健康有显著的正相关（r=0.340，p<0.001；r=0.237，p<0.001），表明 HLS-SF9 和 HLS-SF4 的实证效度良好。此

外，完整版健康素养与领悟社会支持有显著的正相关（r=0361，p<0.001），与家庭健康有显著的正相关（r=0.329，p<0.001），低于 HLS-SF9 与其他量表的相关性，说明 HLS-SF9 的灵敏度更高。

表 5 健康素养量表和两个简版量表针对其他量表的实证效度（n=3370）

相关结构	健康素养量表		
	（CTT）HLS-SF9	（Mokken）HLS-SF4	HLS-SF12
PSSS	0.367**	0.292**	0.361**
FHS-SF	0.340**	0.237**	0.329**

注：**指 p<0.001

2.4.4 测量内容一致性

基于数据集 2，本研究分析了两个简化版健康素养的效标效度。HLS-SF9 对 HLS-SF12 的效标效度的 ICC（95%CI）为 0.989（0.988，0.999），HLS-SF4 效标效度的 ICC（95%CI）为 0.892（0.886，0.899），表明效标效度良好及以上，即两个精简后的量表与完整版量表所测量的内容有高度一致性。

3 讨论

3.1 量表简化过程和结果的合理性

HLS-SF12 运用主成分分析法进行简化，在亚洲六个国家地区的验证结果显示[10]，较好的反映了 HL 的理论结构，并在健康素养水平亚组人群间存在有效差异，校标检验显示 HLS-SF12 对亚洲健康素养量表 HLS-EU-Q47 的解释有效性高于欧洲地区量表 HLS-EU-Q16。HLS-SF12 量表已被国外学者应用于普通门诊、骨科及中医科等科室患者的研究[39]，以及在越南农民、新冠期间卫生工作者及门诊患者等人群中得到验证应用[40-42]，均表明该量表具有良好的信效度，并且在跨文化背景、地域差异以及社会群体差异等方面具有普适性，可作为应用多群体健康素养的有效衡量工具。

以往简化的研究中常用经典测量理论和 Mokken 模型这两个理论模型。一方面，CTT 是历史悠久、发展时间长、应用最广、最为人们熟知的一种心理测量学理论[43]。它将测验观察分数表示为真分数和误差分数之和，并且在其假设的基础上，经过几十年的实践，从理论上推导出包括信度、效度、条目难度和区分度等十几个参数的计算公式，建立了完善测验方法体系，明确了测验标准化程序，使整个测验过程更加客观、科学。除此之外，CTT 拥有一套较为易懂的数学模型、参数概念和估计方法，提倡的标准化技术能有效控制测量过程中产生的误差，更重要的是，它的理论和方法体系相对完整，前提假设比较弱，很容易为实际工作所满足[44]。例如，于斌斌[45]等人采用了基于经典测量理论的极端值法、相关系数法和项目-总体相关系数法将批判思维倾向量表简化为 28 个项目，且信效度

检验结果显示简化版量表可用性强。另一方面，IRT 的测验模型也被证实具有更多的优点，它采用非线性模型，建立了被试对项目的反应与其潜在特质之间的非线性关系，这一点更符合实践领域中的施测情况[46]。而 Mokken 模型是非参数项目反应理论模型的一种，具有非参数的特性，同样服从 IRT 的基本原则，可以弥补参数项目反应理论模型的不足[47]。对所有项目进行 Mokken 模型分析后可以将不符合理论假设的项目加以删除或修改，进一步提高量表的质量[48]，武运筹、王飞等人[49-50]使用 Mokken 模型分别简化了领悟社会支持量表和新的一般自我效能感量表，信效度良好，可应用于实践领域。

在简化前，我们使用验证性因素分析对数据集 1 中原量表的结构效度进行验证，结果表明健康素养量表（HLS-SF12）的结构效度良好。在基于经典测量理论的量表简化过程中，我们还根据五种常用于项目分析的经典测量理论的方法对健康素养量表进行精简，其中每个维度中有一个条目在项目间残差相关法中考虑删除，将其删除后形成一个 9 条目的简版量表（HLS-SF9）。

在根据非参数项目反应理论进行量表简化的过程中，我们对完整版量表进行 Mokken 分析。首先对完整版量表进行分析，单维性检验表明完整版量表只能得到一个维度，同质性系数良好，但在局部独立性检验中，依次删除了具有条件关联的 8 个条目，再进行单调性检验，没有删除多余的项目，最终得到了一个 4 条目的简版量表（HLS-SF4）。

以经典测量理论和 Mokken 模型为基础开展简化工作，得到了两个精简后的版本 HLS-SF9 和 HLS-SF4，利用数据集 2 的数据进行天花板和地板效应分析，结果表明两个效应都低，可以较好区分高低分被试，有利于进行进一步的信度分析。信度检验表明，两个简版量表信度良好。然而，HLS-SF9 的各信度优于 HLS-SF4。在结构效度检验中首先进行了 Bartlett 球形检验和 KMO 度量，确定 HLS-SF4 可以进行因子分析，然后采用探索性因素分析验证 HLS-SF4 的简化的结构，经过主成分分析提取出一个特征根大于 1 的公因子，结果符合 Mokken 模型分析中 AISP 所检验的量表单维结构的前提假设；对 HLS-SF9 则采用验证性因素分析，结果显示其三维模型的各项适配指标均为优秀，说明 9 条目的简版量表被划分为三个维度是合理的。值得一提的是，两个简版量表是采用两种不同的结构效度验证方式检验量表结构，这有两个原因：一是利用经典测量理论简化条目时，并未涉及到维度的增减，维度确定且与原量表保持一致；二是利用 Mokken 模型简化条目时，打破了原有的维度，合并成单一维度，且简化后保留的条目只涉及前两个维度。实证效度检验的结果显示，HLS-SF9 与其他量表的相关性高于原量表以及 HLS-SF4 与其他量表的相关性，说明 HLS-SF9 的灵敏度更高。测量两个精简后的量表与完整版量表的 ICC 指数，结果表明所要测量的内容有高度一致性，但 HLS-SF9 高于 HLS-SF4。综上所述，HLS-SF9 条目保留原量表因子结构，信度、灵敏度更优，和原量表的测量内容一致性更高，而 HLS-SF4 条目少，各项因子载荷量更高，因此保留两个精简版本均具有合理性。

在传染性或非传染性疾病的预防与控制中，健康素养是不容忽视的重要因素，比如在新型冠状肺炎的信息疫情中，健康素养成为公众辨识“谣言”的关键工具[51]。本研究使用全国范围的大样本研究数据，尽可能降低地域差异带来的偏倚，

以提高研究的可推广性[52]，并将数据随机生成两个样本群，相互验证研究结果。精简后的健康素养量表条目相较于国内的常用的评估问卷，条目数量少、作答时间短、作答难度较低，更适宜测评全年龄段人群亦或在综合性问卷中使用。

3.2 局限及未来研究

本研究分别采用两种理论为基础，皆严格遵守了量表简化的理论原则，但仍存在一定局限性。例如本研究选择全国大样本横截面研究为数据来源，未来需要在纵向研究中，做简化版量表的有效性与稳定性验证工作；在多领域实际运用中，也需要更多的适应性验证与调整。

4 结论

本研究运用经典测量理论以及 Mokken 模型筛选 HLS-SF12 条目，经过简化的 9 条目与 4 条目的健康素养量表，在我国公众群体中具有良好的信效度，可以作为测评我国全人群健康素养的可靠且精简的工具。

参考文献：

[1] Ad Hoc Committee on Health Literacy for the Council on Scientific Affairs, American Medical Association. Health Literacy: Report of the Council on Scientific Affairs[J]. *Journal of the American Medical Association*, 1999, 281:552-557. doi:10.1001/jama.281.6.552

[2] 别如娥.基于突发公共卫生事件的居民健康素养评价指标体系研究[D].南京:南京中医药大学,2012.

[3] Sheiham A . Commission on Social Determinants of Health (CSDH) Closing the Gap in a Generation: Health Equity Through Action on Social Determinants of Health[. *Community Dental Health*, 2009, 26(1):2-3.

[4] Muir KW, Santiago-Turla C, Stinnett SS, et al. Health literacy and adherence to glaucoma therapy[J]. *Am J Ophthalmol*. 2006;142(2):223-226. doi:10.1016/j.ajo.2006.03.018

[5] Baker DW, Gazmararian JA, Williams MV, et al. Functional health literacy and the risk of hospital admission among Medicare managed care enrollees[J]. *Am J Public Health*. 2002;92(8):1278-1283. doi:10.2105/ajph.92.8.1278

[6] Sudore RL, Yaffe K, Satterfield S, et al. Limited literacy and mortality in the elderly: the health, aging, and body composition study[J]. *J Gen Intern Med*. 2006;21(8):806-812. doi:10.1111/j.1525-1497.2006.00539.x

[7] Nie X, Li Y, Li C, Wu J, Li L. The Association Between Health Literacy and Self-rated Health Among Residents of China Aged 15-69 Years[J]. *Am J Prev Med*. 2021;60(4):569-578. doi:10.1016/j.amepre.2020.05.032

[8] 李婷,易巧云,孙玫.健康素养评估工具的研究进展[J].解放军护理杂志,2015,32(18):29-32.

[9] 欧阳煜,王秀华,杨琛,谭哲煜.老年人健康素养评估工具研究进展[J].解放军护理杂志,2018,35(02):39-43+48.

[10] Duong T V , Aringazina A , Kayupova G , et al. Development and Validation of a New Short-Form Health Literacy Instrument (HLS-SF12) for the General Public in Six Asian Countries[J]. *HLRP Health Literacy Research and Practice*, 2019, 3 (2) : 91-102. doi: 10.3928/24748307-20190225-01.
[11] 欧光忠,张山鹰,管纪惠.应用德尔菲法构建糖尿病患者健康信息素养评价指标体系[J].中国健康教育,2014,30(02):107-110.
[12] 王高玲,蒋欣静,张怡青.慢性病患者健康素养评价指标体系 Delphi 法构建[J].中国公共卫生,2018,34(01):71-74.
[13] 晋菲斐. 中国农村居民健康素养量表研制与评价[D].中国疾病预防控制中心,2018.
[14] 路桃影,吴大嵘.简短版量表研究中条目筛选方法概述.乌鲁木齐:第七届中医/中西医结合循证医学方法研讨会, 2013:96-102.
[15] Cawthon C , Mion L C , Willens D E , et al. Implementing Routine Health Literacy Assessment in Hospital and Primary Care Patients[J]. *Joint Commission Journal on Quality & Patient Safety*, 2014, 40(2):68-76.
[16] Zimet GD, Powell SS, Farley GK, Werkman S, Berkoff KA. Psychometric characteristics of the Multidimensional Scale of Perceived Social Support[J]. *J Pers Assess*. 1990;55(3-4):610-617. doi:10.1080/00223891.1990.9674095
[17] Crandall A, Weiss-Laxer NS, Broadbent E, et al. The Family Health Scale: Reliability and Validity of a Short- and Long-Form[J]. *Front Public Health*. 2020;8:587125. Published 2020 Nov 20. doi:10.3389/fpubh.2020.587125
[18] Wang F, Wu Y, Sun X, et al. Reliability and validity of the Chinese version of a short form of the family health scale[J]. *BMC Prim Care*. 2022;23(1):108. Published 2022 May 6. doi:10.1186/s12875-022-01702-1
[19] Stevens ML, Crowley P, Garde AH, Mortensen OS, Nygård CH, Holtermann A. Validation of a Short-Form Version of the Danish Need for Recovery Scale against the Full Scale[J]. *Int J Environ Res Public Health*. 2019;16(13):2334. Published 2019 Jul 2. doi:10.3390/ijerph16132334
[20] Günüç S, Odabasi H F, Kuzu A. Developing an effective lifelong learning scale (ELLS): Study of validity & reliability[J]. Egitim ve Bilim, 2014, 39(171).
[21] 柴晓运,李晓燕,曹娟,林丹华.中国青少年积极发展量表（简版）的编制：基于一个大样本数据[J].心理与行为研究,2020,18(05):631-637.
[22] Olatunji BO, Williams NL, Tolin DF, et al. The Disgust Scale: item analysis, factor structure, and suggestions for refinement[J]. *Psychol Assess*. 2007;19(3):281-297. doi:10.1037/1040-3590.19.3.281
[23] Jaarsma T, Arestedt KF, Mårtensson J, Dracup K, Strömberg A. The European Heart Failure Self-care Behaviour scale revised into a nine-item scale (EHFScB-9): a reliable and valid international instrument[J]. *Eur J Heart Fail*. 2009;11(1):99-105. doi:10.1093/eurjhf/hfn007
[24] Coşkun Y D, Demirel M. Lifelong learning tendency scale: the study of validity and reliability[J]. Procedia-Social and Behavioral Sciences, 2010, 5: 2343-2350.
[25] Mokken, R. J. (1971). A theory and procedure of scale analysis. The Hague, The Netherlands: Mouton.
[26] Junker B W, Sijtsma K. Cognitive assessment models with few assumptions, and connections with nonparametric item response theory[J]. Applied Psychological Measurement, 2001, 25(3): 258-272.doi:10.1177/01466210122032064.

[27] Hemker B T, Sijtsma K, Molenaar I W. Selection of unidimensional scales from a multidimensional item bank in the polytomous Mokken I RT model[J]. Applied Psychological Measurement, 1995, 19(4): 337-352.
[28] 张军.非参数项目反应理论在维度分析中的运用及评价[J].心理学探新,2010,30(03):80-83.
[29] Straat J H, van der Ark L A, Sijtsma K. Using conditional association to identify locally independent item sets[J]. Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, 2016, 12(4): 117.
[30] Sijtsma K, Meijer R R, van der Ark L A. Mokken scale analysis as time goes by: An update for scaling practitioners[J]. Personality and Individual Differences, 2011, 50(1): 31-37.
[31] 彭旺. 高职生自我批评反刍的测量、现状及其与拖延的关系研究[D].贵州师范大学,2020.DOI:10.27048/d.cnki.ggzsu.2020.001040.
[32] Molenaar, I. W., & Sijtsma, K. (2000). MPS5 for Windows. A program for Mokken scale analysis for polytomous items. iec ProGAMMA.
[33] Terwee C B, Bot S D M, de Boer M R, et al. Quality criteria were proposed for measurement properties of health status questionnaires[J]. Journal of clinical epidemiology, 2007, 60(1): 34-42.doi:10.1016/j.jclinepi.2006.03.012
[34] Messner E M, Terhorst Y, Barke A, et al. The German version of the Mobile App Rating Scale (MARS-G): development and validation study[J]. JMIR mHealth and uHealth, 2020, 8(3): e14479.
[35] Koo T K, Li M Y. A guideline of selecting and reporting intraclass correlation coefficients for reliability research[J]. Journal of chiropractic medicine, 2016, 15(2): 155-163.
[36] 张静,齐玉龙,马嫣,陈玮.父母教养方式、领悟社会支持与大学生健康素养的关系[J].蚌埠医学院学报,2017,42(04):483-486+490.
[37] 张微,刘瑶,赵雅宁,郝红梅,张源,顾小颖,郑光磊,彭通,李雪梅.家庭功能和社区卫生服务对社区居民传染病健康素养影响的研究[J].护理管理杂志,2022,22(05):331-335.
[38] 李浩淼,孙菊,姚强,吴一波.家庭健康: 概念、测度与中国现状[J].中国卫生政策研究,2022,15(04):1-7.
[39] Duong TV, Chang PW, Yang SH, et al. A New Comprehensive Short-form Health Literacy Survey Tool for Patients in General[J]. Asian Nurs Res (Korean Soc Nurs Sci). 2017, 11(1):30-35. doi:10.1016/j.anr.2017.02.001.
[40] Duong T V , Nguyen T , Pham K M , et al. Validation of the Short-Form Health Literacy Questionnaire (HLS-SF12) and Its Determinants among People Living in Rural Areas in Vietnam[J]. International Journal of Environmental Research and Public Health, 2019, 16 (18) : 3346. doi: 10.3390/ijerph16183346.
[41] Tran TV, Nguyen HC, Pham LV, et al. Impacts and interactions of COVID-19 response involvement, health-related behaviours, health literacy on anxiety, depression and health-related quality of life among healthcare workers: a cross-sectional study [J]. BMJ Open, 2020, 10 (12): 041394. doi:10.1136/bmjopen-2020-041394.
[42] Pham KM, Pham LV, Phan DT, et al. Healthy Dietary Intake Behavior Potentially Modifies the Negative Effect of COVID-19 Lockdown on Depression: A Hospital and Health Center Survey [J]. Front Nutr, 2020, 7 : 581043.2020, 7:581043. doi:10.3389/fnut.2020.581043.
[43] 孙晓敏,关丹丹.经典测量理论与项目反应理论的比较研究[J].中国考试(研究版),2009(09):10-17.DOI:10.19360/j.cnki.11-3303/g4.2009.09.003.
[44] 杨志明,张雷著, 测评的概化理论及其应用[M]. 教育科学出版社 , 2003

[45] 于斌斌,于良芝.面向 LIS 研究与实践的批判思维倾向量表：对 CTDI-CV 进行简化的可行性检验[J].图书与情报,2020(04):21-27.
[46] Embretson S E, Reise S P. Item response theory[M]. Psychology Press, 2013.
[47] Van der Ark L A. Mokken scale analysis in R[J]. Journal of statistical software, 2007, 20: 1-19.
[48] 袁淑莉,何壮.非参数项目反应理论模型——Mokken 模型[J].贵阳学院学报(自然科学版),2020,15(04):101-106.DOI:10.16856/j.cnki.52-1142/n.2020.04.024.
[49] Wang F, Chen K, Du Z, et al. Reliability and validity analysis and Mokken model of New General Self-Efficacy Scale-Short Form (NGSES-SF). doi:10.31234/osf.io/r7aj3.
[50] Wu Y, Tang J, Du Z, et al. Development of a Short Version of the Perceived Social Support Scale: Based on Classical Test Theory and Item Response Theory. doi:10.31234/osf.io/d95zg.
[51] Paakkari L, Okan O. COVID-19: health literacy is an underestimated problem[J]. *Lancet Public Health*. 2020;5(5):e249-e250. doi:10.1016/S2468-2667(20)30086-4.
[52] Shin H R, Choi E Y, Kim S K, et al. Health literacy and frailty in community-dwelling older adults: Evidence from a nationwide cohort study in South Korea[J]. International Journal of Environmental Research and Public Health, 2021, 18(15): 7918.

作者贡献声明：

孙小楠，王飞：提出研究思路，设计研究方案；

孙小楠，陈珂，武运筹：研究过程的实施；

孙小楠，陈珂，武运筹：采集、清洗和分析数据；

孙小楠，陈珂：论文起草；

孙小楠，陈珂，汤靖琪：论文最终版本修订；

贺苗，吴一波，孙昕霙：参与并监督整个研究过程。